Enstore with Chimera namespace provider

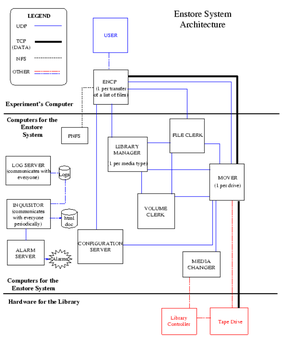
D.Litvintsev, A.Moibenko, G.Oleynik, M.Zalokar
(*Fermi National Accelerator Laboratory*)

Enstore is Hierarchical Storage Management system developed and operated by Fermilab. It provides seamless access to the data stored on permanent media by client applications distributed across IP network.

Enstore is a client-server application.

Client side:
- Encp client provides “cp” like functionality to retrieve/store files from/to tape.

The server side is a multicomponent ensemble of distributed servers that provide:
- Hierarchical view of files stored on tape presented to user as it were a Unix file system. PNFS namespace provider developed by DESY is used.
- Management of user files
- Distributed access to tape drives
- Interface to Robotic Tape Libraries
- Resource management (tapes, drives)
- Tape allocation accounting per storage group, media type
- Self-monitoring, error-reporting and alarm services

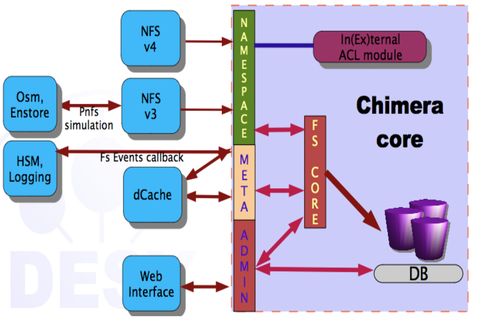


3 Enstore instances
40 PB on tape

Namespace provider functions:
- Unique file ID
- Path to ID mapping
- Mechanism to store file metadata
- Directory tags inherited by subdirectories
- Callbacks on FS events

Enstore uses PNFS implementation of namespace provider developed in 1997 by DESY. NFSv2 on top of DB.
Limitations:
- Max file size is 2 GB
- Metadata access only through NFS
- Metadata stored at BLOBS
- No ACLs, no security

PNFS is de-supported product.



Chimera is the next generation namespace provider implementation provided by DESY

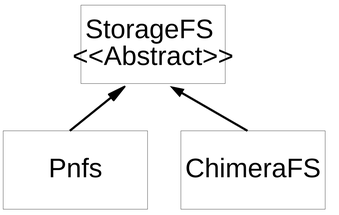
High performance replacement for PNFS
- Built on top of Relational DB, allowing efficient metadata querying
- Well defined API for metadata and namespace operations and admin interface
- Platform independent

- Plugin interface for permission handler
- NFS version supported:
  - v2(legacy), v3(legacy) no 2GB size limit
  - v4 with GSS authentication
  - v4.1 with parallel POSIX I/O. A real filesystem

PNFS->Chimera replacement entails adaptation of encp (Enstore client) only

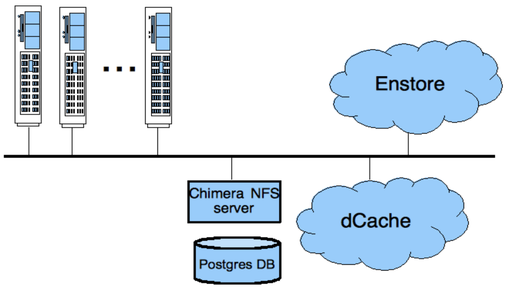
Encp uses factory method to instantiate concrete implementation of StorageFS class at runtime based on top directory tags.

Chimera support is available in encp v3_10e



Acceptance test

100 client nodes



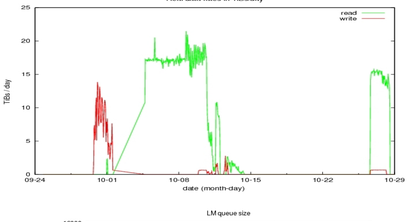
- Attached to SL8500
- 2 LMs
- 6 LTO4 tape drives
- Chimera mounted @ 100 client nodes
- 130 encps/client

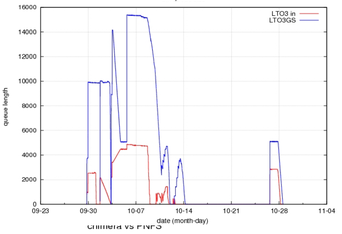
- End-to-end tests with dCache :
- 130 dccp r/w per node

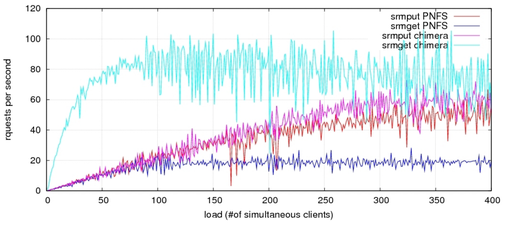
#transfers	TiBs	Enstore/dCache	r/w
83198	161.5	Enstore	read
23749	25.4	Enstore	write
1215	2.3	dCache	read
9869	1.4	dCache	write

Load similar or exceeding production system load:
- Up to 18K LM queue size

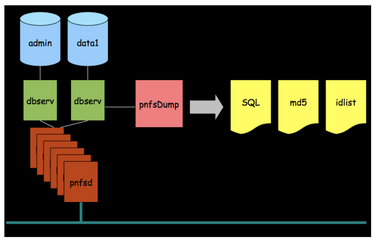
No Errors observed







Migration, Deployment in production



Migration involves:
- PnfsDump
- SQL Injection
- Location info population
- Companion DB migration
- md5sum verification

pnfsDump	6h52m
SQL import	9h47m
enstore2chimera	1h8m
import of companion	17m
md5sum verification	113h24m